



Московский государственный технический университет
имени Н.Э. Баумана

Методические указания

Н.С. Полякова, Г.С. Дерябина, Х.Р. Федорчук

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Издательство МГТУ им. Н.Э. Баумана

Московский государственный технический университет
имени Н.Э. Баумана

Н.С. Полякова, Г.С. Дерябина, Х.Р. Федорчук

МАТЕМАТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
И ПЛАНИРОВАНИЕ
ЭКСПЕРИМЕНТА

*Методические указания
к выполнению домашнего задания*

Москва
Издательство МГТУ им. Н.Э. Баумана
2010

УДК 517+519.2
ББК 22.16+22.17
П54

Рецензент *А.А. Грешилов*

Полякова Н.С.
П54 Математическое моделирование и планирование эксперимента: метод. указания к выполнению домашнего задания / Н.С. Полякова, Г.С. Дерябина, Х.Р. Федорчук. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2010. – 33, [3] с.

Изложены требования к построению математических моделей. Рассмотрены свойства математических моделей, метод наименьших квадратов для однократных и повторных наблюдений, а также методика обработки данных эксперимента.

Для студентов старших курсов.

УДК 517+519.2
ББК 22.16+22.17

ВВЕДЕНИЕ

Математической моделью объекта называется совокупность математических зависимостей, описывающих его функционирование.

Реальные объекты достаточно сложны, поэтому при математическом описании их неизбежно приходится упрощать, огрублять, пренебрегая второстепенными факторами и обращая внимание на существенные. Успех или неудача исследователя во многом предопределяется выбором способа описания объекта, вида математической модели. «Математическая модель — это вопрос, задаваемый исследователем природе. Он, как и всякий вопрос, содержит как утверждающую часть — наши знания о явлении, — так и вопрошающую часть — то, что мы хотим узнать» [1].

Математические модели можно получать теоретическими и экспериментальными методами. Модели, построенные теоретическими методами, называют аналитическими, модели, полученные по результатам обработки экспериментальных данных, называют эмпирическими. Но наиболее эффективно сочетание аналитических методов с экспериментальными, поскольку математическое описание объекта, полученное аналитически, всегда содержит константы, которые определяются по результатам эксперимента.

Статистическую обработку результатов эксперимента при получении эмпирических моделей часто проводят методами регрессионного анализа. Такие эмпирические модели называют регрессионными. Рекомендациями по составлению планов и получением математических моделей по результатам их реализации занимается специальная дисциплина — планирование эксперимента.

В зависимости от способа представления информации различают детерминированные и вероятностные (или стохастические)

математические модели. Вероятностные модели содержат случайные параметры, поэтому результат расчета по ним — это либо вероятность наступления определенного события, либо статистическая оценка некоторой случайной величины. Все регрессионные модели также являются вероятностными, поскольку для них выходной величиной является статистическая оценка условного математического ожидания некоторого параметра.

В отличие от вероятностной детерминированная модель однозначно предсказывает значение выходной величины при заданных значениях входных параметров.

Область основного применения вероятностных моделей — описание объектов в условиях неопределенности, т. е. при отсутствии некоторых сведений об условиях функционирования.

Использование математической модели может быть эффективным, если она обладает следующими свойствами: полнота, точность, адекватность, устойчивость, продуктивность, наглядность.

Полнота математической модели — это ее способность отразить в достаточно полной мере те характеристики объекта, которые интересуют исследователя.

Точность математической модели — это оценка относительной погрешности найденных с помощью математической модели значений выходных параметров. Она дает возможность обеспечить приемлемое совпадение реальных и найденных с помощью модели значений выходных параметров изучаемого объекта.

Адекватность математической модели — это ее способность дать правильное качественное и достаточно точное количественное описание именно тех характеристик объекта, которые важны в данном конкретном случае.

Устойчивость математической модели означает ее способность нивелировать погрешности в исходных данных, не допускать их чрезмерного влияния на результаты расчетов по модели. Причинами низкой устойчивости могут быть деления на слишком малые по модулю величины, использование слишком быстро растущих функций. В случае системы линейных уравнений устойчивость определяется отношением наибольшего и наименьшего собственных чисел системы. Иногда стремление увеличить полноту математической модели приводит к снижению ее устойчивости вследствие введения дополнительных параметров, известных

с невысокой точностью или входящих в слишком приближенные соотношения.

Продуктивность математической модели связана с тем, что точность выходных параметров модели не должна превосходить точность входных параметров, иначе ее применение для изучения конкретного объекта теряет смысл.

Наглядность математической модели является желательным, но не обязательным свойством.

Математические модели могут быть разделены на структурные и функциональные.

Структурные математические модели отражают структуру устройства изучаемого объекта. Структурные модели также разделяются на топологические и геометрические. Топологические модели отражают состав объекта и связи между его элементами, применяются обычно на начальной стадии исследования объекта. Геометрическая модель включает информацию топологической модели и, кроме того, содержит сведения о форме и размерах элементов объекта, их взаимном расположении, что в свою очередь может задаваться с помощью уравнений поверхностей и различных алгебраических соотношений.

Функциональные математические модели могут отражать физические, химические или информационные процессы, протекающие в объекте. Однако функционирование сложных систем нередко удается описать лишь с помощью совокупности их реакций на входные воздействия. Такую разновидность функциональных моделей относят к типу черного ящика и называют **имитационной математической моделью**, имея в виду, что она лишь имитирует внешние проявления функционирования объекта, не описывая существа протекающих в нем процессов.

1. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Статистическая зависимость случайной величины Y от другой величины X в среднем описывается функцией регрессии: $\varphi_Y(x) = M(Y|X = x)$, здесь M — математическое ожидание. Различают несколько видов регрессии: регрессию Y на X и регрессию X на Y . В первом случае считается, что случайной величиной является Y , а аргументы X являются детерминированными или известны с погрешностью много меньшей, чем погрешность величины Y .

В регрессии X на Y имеет место обратная ситуация. Если погрешности величин соизмеримы, т. е. ни одной погрешностью пренебречь нельзя, то применяются методы конфлюэнтного анализа, учитывающие одновременно погрешности величин X и величин Y . Далее рассматриваются только регрессии Y на X .

Если теоретическая функция регрессии $y = \varphi(x)$ неизвестна, ее можно найти по эмпирическим данным. Особенно удобно это делать, если известна параметрическая форма: $y = \varphi(x, \theta_1, \dots, \theta_p)$, а еще проще, если эта зависимость линейная:

$$y = \theta_0 + \theta_1 f_1(x) + \dots + \theta_p f_p(x), \quad (1)$$

где f_1, \dots, f_p — известные функции и, возможно, x — многомерная величина, т. е. $x = (x_1, \dots, x_k)$.

Пусть проведено всего N опытов, y_i — результат i -го опыта (эмпирическая величина), тогда $y_i = \theta_0 + \theta_1 f_1(x) + \dots + \theta_p f_p(x) + \varepsilon_i$, где ε_i — случайная ошибка (погрешность). Относительно ε_i мы предполагаем, что математические ожидания $M\varepsilon_i = 0$, т. е. систематическая ошибка отсутствует, а дисперсии $D\varepsilon_i = \sigma^2$, и что $M(\varepsilon_i \varepsilon_j) = 0, i \neq j$, т. е. случайные ошибки некоррелированы. Обозначим X_{ij} — значение f_j -й функции в i -м опыте. Наблюдаемые величины y_1, \dots, y_N также считаем некоррелированными и

имеющими одинаковую дисперсию σ^2 , т.е. $\text{cov}(y_i, y_j) = \sigma^2$ (при $i = j$) и $\text{cov}(y_i, y_j) = 0$ при $i \neq j$. Следовательно, корреляционная матрица будет иметь вид

$$D\{Y\} = \begin{pmatrix} \text{cov}(y_1, y_1) & \dots & \text{cov}(y_1, y_N) \\ \dots & \dots & \dots \\ \text{cov}(y_N, y_1) & \dots & \text{cov}(y_N, y_N) \end{pmatrix} = \sigma^2 I_N,$$

где σ^2 — неизвестный параметр; I_N — единичная $N \times N$ -матрица. Таким образом, получаем

$$MY = X\Theta; \quad DY = \sigma^2 I_N,$$

где

$$X = \begin{pmatrix} X_{10} & \dots & X_{1p} \\ \dots & \dots & \dots \\ X_{N0} & \dots & X_{Np} \end{pmatrix}$$

— матрица известных коэффициентов; $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ — вектор неизвестных параметров; $Y = (Y_1, \dots, Y_N)^T$ — вектор откликов.

Таким образом, получаем линейную модель Гаусса — Маркова:

$$\begin{cases} Y = X\Theta + E; \\ ME = 0; \\ DE = \sigma^2 I_N, \end{cases}$$

где $E = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$.

Метод наименьших квадратов состоит в минимизации величины

$$R = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \sum_{j=0}^p x_{ij} \theta_j)^2.$$

Необходимое условие экстремума:

$$0 = \partial R / \partial \theta_k = -2 \left(\sum_{i=1}^N (y_i - \sum_{j=0}^p x_{ij} \theta_j) x_{ik} \right),$$

следовательно,

$$\sum_{i=1}^N \sum_{j=0}^p x_{ij} x_{ik} \theta_j = \sum_{i=1}^N x_{ik} y_i, \quad k = 0, 1, \dots, p.$$

Запишем это равенство в матричном виде:

$$X^T X \Theta = X^T Y.$$

Таким образом получили систему нормальных уравнений. Решение этой системы

$$\tilde{\Theta} = (X^T X)^{-1} X^T Y$$

является точечной оценкой метода параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ по методу наименьших квадратов.

Величина

$$R_0 = \sum_{i=1}^N (y_i - \sum_{j=0}^p x_{ij} \theta_j)^2$$

называется остаточной суммой квадратов,

$$\tilde{\sigma}^2 = \frac{R_0}{N - p - 1}$$

является оценкой дисперсии σ^2 .

Замечание. Переход от значений x_{ij} (где i — номер опыта, j — номер фактора) реальных варьируемых факторов и составленной из них матрицы плана

$$\Pi = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots \\ x_{N1} & \dots & x_{Nk} \end{pmatrix},$$

к матрице

$$X = \begin{pmatrix} X_{10} & \dots & X_{1p} \\ \dots & \dots & \dots \\ X_{N0} & \dots & X_{Np} \end{pmatrix}$$

является переходом к матрице базисных функций. Здесь $X_{i0} = 1$, $i = 1, \dots, N$ — значения фиктивного фактора для нахождения константы θ_0 , $X_{ji} = f_i(x_{j1}, \dots, x_{jk})$ — значение i -й функции в j -м опыте.

Пример 1. Имеется один варьируемый фактор X , принимающий значения $-1, 0, 1$, следовательно, $\Pi = (-1, 0, 1)^T$ — матрица плана. $Y = \theta_0 + \theta_1 X + \theta_2 X^2$ — модель (функция отклика), тогда матрица базисных функций будет иметь вид

$$X = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Пример 2. Есть два фактора, матрица плана:

$$\Pi = \begin{pmatrix} 2 & 5 \\ 5 & 0 \\ 4 & 8 \\ 0 & 6 \end{pmatrix},$$

модель: $Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_1 X_2$, следовательно, матрица базисных функций будет иметь вид

$$X = \begin{pmatrix} 1 & 2 & 5 & 10 \\ 1 & 5 & 0 & 0 \\ 1 & 4 & 8 & 32 \\ 1 & 0 & 6 & 0 \end{pmatrix}.$$

Пример 3. Пусть $Y = \theta_0 + \theta_1 X + \theta_2 X^2$, $N = 10$, тогда

$$X = \begin{pmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ \dots\dots\dots \\ 1 & X_{10} & X_{10}^2 \end{pmatrix}$$

— матрица базисных функций;

$$X^T X = \begin{pmatrix} 10 & \sum_{i=1}^{10} X_i & \sum_{i=1}^{10} X_i^2 \\ \sum_{i=1}^{10} X_i & \sum_{i=1}^{10} X_i^2 & \sum_{i=1}^{10} X_i^3 \\ \sum_{i=1}^{10} X_i^2 & \sum_{i=1}^{10} X_i^3 & \sum_{i=1}^{10} X_i^4 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} \sum_{i=1}^{10} y_i \\ \sum_{i=1}^{10} X_i y_i \\ \sum_{i=1}^{10} X_i^2 y_i \end{pmatrix}.$$

Замечание:

$$D(\Theta) = \sigma^2(X^T X)^{-1} \quad (2)$$

— матрица ковариаций для Θ , т. е. дисперсия $D\theta_i = \sigma^2 c_{ii}$, а $\text{cov}(\theta_i, \theta_j) = \sigma^2 c_{ij}$, где

$$\begin{pmatrix} c_{00} & \dots & c_{0p} \\ \dots & \dots & \dots \\ c_{p0} & \dots & c_{pp} \end{pmatrix} = (X^T X)^{-1}.$$

Если план насыщенный, т. е. число опытов равно числу неизвестных параметров, то оценку σ^2 найти невозможно.

2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ ПРИ ПОВТОРНЫХ НАБЛЮДЕНИЯХ

В связи с тем, что исход опыта может зависеть не только от рассматриваемых факторов, но и от некоторых случайных составляющих, которые хоть и незначительно, но оказывают влияние на результат, имеет смысл проводить так называемые повторные наблюдения, т. е. проводить эксперимент несколько раз с одними и теми же значениями рассматриваемых факторов.

Пусть в точке $\tilde{x}_1 = (x_{11}, x_{21}, \dots, x_{k1})$ наблюдение проводилось m_1 раз, в точке $\tilde{x}_2 = (x_{12}, \dots, x_{k2})$ — m_2 раз, ..., в точке $\tilde{x}_n = (x_{1n}, \dots, x_{kn})$ — m_n раз. Матрицу

$$\tilde{\Pi} = \begin{pmatrix} x_{11} & \dots & x_{k1} \\ \dots & \dots & \dots \\ x_{1n} & \dots & x_{kn} \end{pmatrix}$$

будем называть матрицей спектра плана, тогда матрица плана Π будет иметь вид

$$\Pi = \begin{pmatrix} x_{11} \dots x_{k1} \\ \dots \dots \dots (m_1 \text{ раз}) \\ x_{11} \dots x_{k1} \\ x_{12} \dots x_{k2} \\ \dots \dots \dots (m_2 \text{ раз}) \\ x_{12} \dots x_{k2} \\ \dots \dots \dots \\ x_{1n} \dots x_{kn} \\ \dots \dots \dots (m_n \text{ раз}) \\ x_{1n} \dots x_{kn} \end{pmatrix}.$$

Соответственно меняется и матрица базисных функций;

$$\tilde{X} = \begin{pmatrix} X_{01} & X_{11} & \dots & X_{p1} \\ \dots \dots \dots \\ X_{0n} & X_{1n} & \dots & X_{pn} \end{pmatrix}$$

— матрица спектра базисных функций;

$$X = \begin{pmatrix} X_{01}X_{11} \dots X_{p1} \\ X_{01}X_{11} \dots X_{p1} \\ \dots \dots \dots (m_1 \text{ раз}) \\ X_{01}X_{11} \dots X_{p1} \\ X_{02}X_{12} \dots X_{p2} \\ \dots \dots \dots (m_2 \text{ раз}) \\ X_{02}X_{12} \dots X_{p2} \\ \dots \dots \dots \\ X_{0n}X_{1n} \dots X_{pn} \\ \dots \dots \dots (m_n \text{ раз}) \\ X_{0n}X_{1n} \dots X_{pn} \end{pmatrix}$$

— матрица базисных функций, следовательно,

$$X^T X = \begin{pmatrix} \sum_{i=1}^n X_{0i}^2 m_i & \sum_{i=1}^n X_{0i}X_{1i} m_i & \dots & \sum_{i=1}^n X_{0i}X_{pi} m_i \\ \dots \dots \dots \\ \sum_{i=1}^n X_{pi}X_{0i} m_i & \sum_{i=1}^n X_{pi}X_{1i} m_i & \dots & \sum_{i=1}^n X_{pi}^2 m_i \end{pmatrix}.$$

Пусть

$$V = \begin{pmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & m_n \end{pmatrix}$$

— матрица дублирований, тогда общее число опытов с учетом дублирований $N = m_1 + \dots + m_n$ является следом матрицы V и, следовательно, $X^T X = \tilde{X}^T V \tilde{X}$.

Определение. Если $m_1 = m_2 = \dots = m_n = m$, то дублирование называется равномерным.

Если дублирование равномерное, то $\tilde{X}^T V \tilde{X} = m \tilde{X}^T \tilde{X}$, следовательно, $(X^T X)^{-1} = (\tilde{X}^T \tilde{X})^{-1}/m$. Матрица ковариаций оценок $\tilde{\Theta}$ в случае неравномерного дублирования имеет вид $D\{\tilde{\Theta}\} = \sigma^2(\tilde{X}^T V \tilde{X})^{-1}$. При равномерном дублировании получаем $D\{\tilde{\Theta}\} = \sigma^2(\tilde{X}^T \tilde{X})^{-1}/m$.

Конечно, равномерное дублирование — это очень удобно, однако не всегда осуществимо вследствие сложности эксперимента, необходимости отбрасывания грубых наблюдений и т. п.

Рассмотрим, как можно преобразовать $X^T Y$ при дублировании:

$$\begin{aligned} X^T Y &= \begin{pmatrix} \sum_{i=1}^n X_{0i}(y_{i1} + \dots + y_{im_i}) \\ \dots \\ \sum_{i=1}^n X_{pi}(y_{i1} + \dots + y_{im_i}) \end{pmatrix} = \\ &= \begin{pmatrix} \sum_{i=1}^n m_i X_{0i}(y_{i1} + \dots + y_{im_i})/m_i \\ \dots \\ \sum_{i=1}^n m_i X_{pi}(y_{i1} + \dots + y_{im_i})/m_i \end{pmatrix}. \end{aligned}$$

Обозначим $\tilde{y}_i = (y_{i1} + \dots + y_{im_i})/m_i$ среднее для результатов повторных наблюдений в точке плана, которая соответствует строке матрицы базисных функций $\tilde{X}^{(i)} = (X_{0i}, \dots, X_{pi})$ и $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_n)^T$ — матрице-столбцу средних значений выходной

величины по дублированиям, тогда $X^T Y = (\tilde{X}^T V) \tilde{Y}$ и окончательно получаем $\tilde{\Theta} = (\tilde{X}^T V \tilde{X})^{-1} \tilde{X}^T V \tilde{Y}$. Тогда при равномерном дублировании $\tilde{\Theta} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}$.

3. ОТБРАСЫВАНИЕ ГРУБЫХ НАБЛЮДЕНИЙ

Прежде чем перейти к расчетам коэффициентов модели по методу наименьших квадратов (МНК), надо провести анализ данных: отбросить явные промахи — грубые (ошибочные) наблюдения.

Из i -й серии дублированных опытов временно исключают сомнительный результат y_{ij} , а по оставшимся данным рассчитывают выборочное среднее \tilde{y}_i и оценку дисперсии s^2 . Далее вычисляют величину $t_{\text{расч}}$ по формуле $t_{\text{расч}} = |y_{ij} - \tilde{y}_i|/s$. По уровню значимости q и числу степеней свободы f , связанному с вычислением s^2 (так как число дублей равно m_i и один элемент y_{ij} исключили, то $f = (m_i - 1) - 1 = m_i - 2$), находят $t_{\text{табл}}$ по табл. П1 приложения. Если $t_{\text{расч}} > t_{\text{табл}}$, то элемент y_{ij} окончательно отбрасывается как грубое наблюдение.

Если сомнение вызывают два или три элемента, то временно отбрасывают все три элемента и исследование начинают с того элемента, который ближе к выборочному среднему. Если этот элемент оказался промахом, то остальные являются более грубыми наблюдениями. Если же наименее сомнительный элемент не оказался промахом, то его присоединяют к выборке и исследуют следующий сомнительный элемент и т. д.

4. СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

После получения уравнения регрессии (1), приступают к его статистическому анализу. Основными задачами анализа являются оценка значимости коэффициентов регрессии и проверка адекватности математической модели (1).

Необходимыми предпосылками статистического анализа являются: а) нормальность распределения выходной величины; б) однородность дисперсий дублированных опытов.

Проверка нормальности распределения осуществляется несколькими способами. Один из наиболее строгих способов — проверка по критерию χ^2 . Для этого необходимо иметь выборку достаточно большого объема: от 50 до 150 наблюдений. Описание порядка действий можно найти в работах [2, 3].

Для выборок небольшого объема гораздо удобнее и теоретически более оправданно использование критерия Колмогорова.

Порядок проверки по критерию Колмогорова следующий:

1) найти оценки для математического ожидания a и среднего квадратичного σ . Для этого можно использовать выборочное среднее

$$\tilde{a} = \tilde{y} = (y_1 + \dots + y_n)/n$$

и исправленную выборочную дисперсию

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \tilde{y})^2}{n - 1};$$

2) построить вариационный ряд по выборке:

$$\{y_1, \dots, y_n\} \rightarrow \{y_1^* \leq y_2^* \leq \dots \leq y_n^*\};$$

3) вычислить значения функции нормального распределения с параметрами (a, σ) в точках вариационного ряда:

$$z_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(y_i^* - \tilde{a})/\sigma} e^{-t^2/2} dt = 0,5 + \Phi\left(\frac{y_i^* - \tilde{a}}{\sigma}\right),$$

где Φ — функция Лапласа, таблицу ее значений можно найти в работе [2—4];

4) вычислить значения критерия $D(\tilde{x}_n)$:

$$d_i = \frac{i}{n} - z_i; \quad d_{i+n} = z_i - \frac{i-1}{n}, \quad i = 1, 2, \dots, n;$$

5) выбрать $d_{\text{расч}} = \max\{|d_i|, 1 \leq i \leq 2n\}$;

6) по уровню значимости q и объему выборки n найти значение в $d_{\text{табл}}$ (табл. 2П приложения);

7) если $d_{\text{расч}} < d_{\text{табл}}$, то гипотеза о нормальности принимается.

Приближенная проверка нормальности распределения может быть выполнена с помощью асимметрии и эксцесса:

$$A = \frac{\sum_{i=1}^n (y_i - \tilde{y})^3}{ns^3}$$

— асимметрия;

$$E = \frac{\sum_{i=1}^n (y_i - \tilde{y})^4}{ns^4} - 3$$

— эксцесс;

$$\sigma_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}$$

— среднее квадратичное асимметрии;

$$\sigma_E = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$$

— среднее квадратичное эксцесса.

Если $\frac{E}{\sigma_E} \geq 3$ или $\frac{A}{\sigma_A} \geq 3$, то гипотеза о нормальности отклоняется.

Если $3 > \frac{E}{\sigma_E} \geq 2$, $3 > \frac{A}{\sigma_A} \geq 2$, то для проверки нормальности следует использовать или критерий χ^2 или критерий Колмогорова.

Метод проверки однородности дисперсий можно найти в работах [2–4]. В данном домашнем задании дублируется только один опыт, поэтому проверка однородности дисперсий невозможна.

Для статистического анализа уравнения регрессии (1) прежде всего необходима дисперсия воспроизводимости $s^2\{y\}$. Пусть n_j — число повторений j -го опыта ($n_j > 1$),

$$\tilde{y}_j = \frac{y_{j1} + \dots + y_{jn_j}}{n_j}$$

— среднее по n_j дублированиям j -го опыта,

$$s_j^2 = \frac{\sum_{k=1}^{n_j} (y_{jk} - \tilde{y}_j)^2}{n_j - 1}$$

— дисперсия воспроизводимости j -го опыта, несмещенная оценка дисперсии j -й серии опытов, тогда

$$s^2\{y\} = \frac{\sum_{j=1}^N s_j^2 f_j}{\sum_{j=1}^N (n_j - 1)} = \frac{\sum_{j=1}^N s_j^2 f_j}{f_y}, \quad (3)$$

где

$$f_y = \sum_{j=1}^N f_j = \sum_{j=1}^N (n_j - 1) \quad (4)$$

— число степеней свободы дисперсии воспроизводимости; $f_j = n_j - 1$ — число степеней свободы дисперсии j -й серии дублирований.

При равномерном дублировании формулы (3), (4) несколько упрощаются. Так как $n_j = n$, то $f_j = n_j - 1 = n - 1$, $f_y = N(n - 1)$,

$$s_j^2 = \frac{\sum_{k=1}^n (y_{jk} - \tilde{y}_j)^2}{n - 1}, \quad s^2\{y\} = \frac{\sum_{j=1}^N s_j^2 (n - 1)}{N(n - 1)} = \frac{\sum_{j=1}^N s_j^2}{N}.$$

Если дублируются не все опыты, то для вычисления дисперсии воспроизводимости берутся только те N_1 опытов, которые дублировались, и для них по описанной выше методике вычисляется $s^2\{y\}$ (недублированные опыты просто не учитываются). Например, пусть n_1 раз дублировался только первый опыт, остальные не дублировались. Тогда

$$s^2\{y\} = \frac{s_1^2 f_1}{f_1} = s_1^2 = \frac{\sum_{i=1}^{n_1} (y_{1i} - \tilde{y}_1)^2}{n_1 - 1},$$

при этом $f_y = f_1 = n_1 - 1$ — число степеней свободы $s^2\{y\}$.

При отсутствии дублированных опытов оценка дисперсии воспроизводимости невозможна, т. е. надо продублировать хотя бы один опыт.

Отметим, что дисперсия воспроизводимости $s^2\{y\}$ является точечной оценкой дисперсии σ^2 , которая присутствует в формуле (2)

для матрицы ковариаций столбца коэффициентов $\tilde{\Theta}$, вычисляемой по МНК.

Для того чтобы выявить незначимые коэффициенты, которые можно в математической модели приравнять к нулю, рассмотрим дисперсию коэффициента θ_i :

$$s^2\{\theta_i\} = c_{ii}s^2\{y\},$$

где c_{ii} — i -й диагональный элемент матрицы $(X^T X)^{-1}$; $s^2\{y\}$ — дисперсия воспроизводимости; тогда

$$s(\theta_i) = \sqrt{s^2\{\theta_i\}}$$

— среднее квадратичное отклонение θ_i .

По уровню значимости q и числу степеней свободы f_y , с которым вычислялась дисперсия воспроизводимости $s^2\{y\}$, находим $t_{\text{табл}}$ по табл. П1 приложения t -критерия.

Если $|\theta_i| \leq t_{\text{табл}}s(\theta_i)$, то θ_i можно отбросить (положить $\theta_i = 0$), т. е. θ_i — незначимый коэффициент. В том случае, если план не ортогональный и, следовательно, матрица базисных функций не ортогональна, то после отбрасывания незначимых коэффициентов необходим пересчет оставшихся коэффициентов по МНК (модель другая).

Если θ_i — значимый коэффициент, то, обозначив через τ_i его истинное значение, получим доверительный интервал для τ_i :

$$\theta_i - t_{\text{табл}}s\{\theta_i\} < \tau_i < \theta_i + t_{\text{табл}}s\{\theta_i\}.$$

5. ПРОВЕРКА АДЕКВАТНОСТИ РЕГРЕССИОННОЙ МОДЕЛИ

Пусть N — число серий опытов, $p + 1$ — число коэффициентов регрессии. Проверка адекватности возможна только при $N > p + 1$ и при наличии оценки дисперсии воспроизводимости.

Порядок проверки адекватности регрессионной модели следующий:

$$1) \text{ вычислить вспомогательную величину } S_a = \sum_{i=1}^N (\tilde{y}_i - \hat{y}_i)^2,$$

где \tilde{y}_i — среднее по i -й серии опытов или просто значение i -го

опыта, если нет дублирований; \hat{y}_i — вычисленное по полученной модели значение выходной величины (отклика);

2) вычислить число степеней свободы дисперсии адекватности $f_a = N - (p + 1)$;

3) вычислить дисперсию адекватности $s_a^2 = S_a / f_a$;

4) вычислить $F_{\text{расч}} = s_a^2 / s^2\{y\}$;

5) по уровню значимости q для f_a в числителе и f_y в знаменателе найти значение $F_{\text{табл}}$ в табл. ПЗ приложения;

6) если $F_{\text{расч}} < F_{\text{табл}}$, то модель адекватна.

В основе рассмотренной проверки адекватности лежит проверка гипотезы об однородности дисперсий адекватности и дисперсии, характеризующей ошибку эксперимента (дисперсию y_i и дисперсию $y_i - \hat{y}_i$).

Кроме проверки адекватности можно оценивать эффективность модели.

При наличии дублирования оценку эффективность модели осуществляют следующим образом:

1) вычисляют дисперсию относительно среднего

$$s_c^2 = \frac{\sum_{j=1}^N \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2}{\sum_{j=1}^N n_j - 1};$$

2) вычисляют остаточную дисперсию:

$$s_{\text{ост}}^2 = \frac{\sum_{j=1}^N \sum_{u=1}^{n_j} (y_{ju} - \hat{y}_j)^2}{\sum_{j=1}^N n_j - (p + 1)};$$

3) находят $F_u = s_c^2 / s_{\text{ост}}^2$.

Если $F_u > 3$, то модель считается эффективной.

6. ДОМАШНЕЕ ЗАДАНИЕ ПО МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ И ПЛАНИРОВАНИЮ ЭКСПЕРИМЕНТА

Требуется:

- 1) провести анализ на наличие грубых наблюдений и отбросить их при необходимости;
- 2) найти дисперсию воспроизводимости;
- 3) проверить нормальность распределения выходной величины (по критерию Колмогорова);
- 4) найти коэффициенты регрессионной модели по МНК;
- 5) оценить значимость коэффициентов регрессионной модели, отбросить незначимые коэффициенты, при необходимости провести повторный расчет;
- 6) провести проверку адекватности.

Исходные данные:

- 1) модель (функция отклика):

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2; \quad (5)$$

- 2) план эксперимента:

Номер опыта	1	2	3	4	5	6	7	8	9
X_1	-1	-1	1	1	-1	1	0	0	0
X_2	-1	1	-1	1	0	0	-1	1	0

Девятый опыт дублируется 10 раз;

- 3) таблица наблюдений в точках плана (в скобках указаны номера дублирования 9-го опыта).

7. РЕШЕНИЕ ТИПОВОГО ПРИМЕРА

В качестве примера рассмотрим решение варианта 24.

Проведем анализ на наличие грубых наблюдений. Результат 14,0 временно исключим как сомнительный, вычислим соответственно среднее по оставшимся дублированиям и выборочную дисперсию:

$$\begin{aligned} \bar{y} &= (13,2 + 13,1 + 12,8 + 13,1 + 12,9 + 12,7 + 13,35 + 12,96 + \\ &+ 12,9)/9 \simeq 13,00; \\ s^2 &= (0,2^2 + 0,1^2 + 0,2^2 + 0,1^2 + 0,1^2 + 0,3^2 + 0,35^2 + 0,04^2 + \\ &+ 0,1^2)/8 \simeq 0,204^2. \end{aligned}$$

Таблиця 1

Номер вари- анта	Номер опыта																	
	1	2	3	4	5	6	7	8	9(1)	9(2)	9(3)	9(4)	9(5)	9(6)	9(7)	9(8)	9(9)	9(10)
1	1,09	1,01	-1,01	7,01	0,01	2,0	-0,10	4,0	1,0	1,01	0,99	1,02	0,94	1,06	0,91	1,50	1,03	0,95
2	3,01	3,1	-1,1	15,1	2,02	6,03	1,04	9,01	4,02	4,01	3,99	4,13	3,94	4,06	4,71	4,03	3,92	3,95
3	7,01	7,02	1,01	25,3	6,02	12,1	4,02	16,4	9,04	9,02	8,99	9,01	8,91	9,5	9,03	9,06	8,97	8,98
4	13,1	13,0	5,01	37,5	12,2	20,3	9,01	25,2	16,1	16,2	15,9	16,0	17,1	15,9	15,8	16,1	16,09	16,07
5	-3,1	21,2	29,3	21,3	9,02	25,4	12,2	20,3	16,2	16,1	15,8	16,3	16,0	14,8	15,6	16,2	16,08	15,92
6	-5,05	13,2	19,1	13,3	4,01	16,3	6,07	12,2	9,09	9,01	9,05	9,07	8,51	8,91	8,92	8,93	7,59	9,04
7	-3,01	-1,0	11,2	7,04	1,01	1,02	2,01	6,03	4,12	4,06	4,04	3,97	3,96	3,98	4,01	4,02	3,91	3,01
8	-3,02	3,01	5,03	3,0	0,02	4,04	0,01	2,02	1,01	1,10	1,02	1,03	0,98	0,97	0,99	1,07	1,54	0,91
9	-7,01	9,2	5,03	21,2	2,01	14,1	-2,0	14,2	7,1	7,12	7,11	7,15	7,8	7,07	6,9	6,87	6,92	6,8
10	-9,2	11,1	7,08	27,3	1,95	18,2	-2,1	17,9	9,09	9,15	9,07	9,13	8,92	8,87	8,2	9,11	8,78	9,1
11	-11,1	12,9	9,05	33,4	2,1	22,3	-1,9	22,2	10,8	10,9	12,1	11,2	10,9	10,8	10,95	11,1	11,2	11,3
12	-13,1	15,1	11,2	39,4	1,98	26,3	-1,95	25,8	12,9	13,1	13,2	13,1	13,3	12,9	12,85	12,8	12,9	14,1
13	1,05	1,03	-1,1	6,97	0,01	2,04	-1,02	4,15	1,06	1,05	0,98	0,96	0,94	1,03	1,09	1,03	0,55	0,95
14	3,1	3,05	-1,1	15,2	2,1	5,95	2,97	6,91	4,03	4,02	3,97	3,95	3,92	4,13	4,98	3,93	3,89	3,96
15	6,9	7,1	1,1	25,4	6,1	12,2	4,09	16,2	9,06	9,05	8,97	8,93	8,92	9,89	9,14	8,96	8,91	8,89
16	13,2	12,9	5,1	36,7	12,1	20,3	9,1	15,9	16,1	16,2	15,8	16,0	17,1	15,8	15,85	16,1	16,2	15,9

Окончание таблицы

Номер вари- анта	Номер опыта																	
	1	2	3	4	5	6	7	8	9(1)	9(2)	9(3)	9(4)	9(5)	9(6)	9(7)	9(8)	9(9)	9(10)
17	-3,1	21,2	29,3	21,1	9,0	25,1	12,2	20,3	16,2	16,1	15,9	15,8	16,3	16,2	15,87	15,1	16,0	15,8
18	-5,2	13,1	19,2	12,9	4,15	16,1	6,05	12,2	9,01	9,02	9,10	9,15	8,89	8,95	9,05	9,95	8,85	8,91
19	-3,1	-1,0	11,2	6,95	1,03	1,01	2,02	6,12	4,07	4,11	4,01	3,97	3,92	4,14	4,03	4,13	4,93	3,89
20	-3,05	2,96	5,03	3,02	0,01	4,04	0,02	2,01	1,02	1,11	1,13	0,89	0,96	0,08	1,05	0,95	0,94	1,08
21	-7,02	9,01	5,02	21,5	2,02	14,2	-2,03	14,1	7,04	7,03	7,10	7,01	6,93	6,89	6,95	7,95	6,91	7,15
22	-9,03	11,0	7,04	27,3	2,01	18,1	-2,02	18,2	9,2	9,1	9,92	9,02	9,11	8,89	8,97	8,87	8,92	9,15
23	-11,1	13,2	9,04	33,2	2,05	22,2	-2,01	22,5	11,3	11,4	10,9	11,5	10,7	11,1	10,8	10,6	9,6	11,2
24	-12,9	15,2	11,1	38,8	2,04	26,3	-2,02	26,3	13,2	13,1	12,8	13,1	12,9	12,7	13,35	14,0	12,96	12,9

Вычислим следующее значение:

$$t_{\text{расч}} = |14,0 - 13,0|/0,204 \simeq 4,90.$$

По уровню значимости $q = 0,05$ и числу степеней свободы $f = 9 - 1 = 8$, связанному с вычислением выборочной дисперсии s^2 , в табл. 1П приложения найдем значение $t_{\text{табл}} = 2,31$, и так как $t_{\text{расч}} > t_{\text{табл}}$, то результат 14,0 отбрасываем.

Так как дублируется только 9-й опыт, то вычисленное значение s^2 , которое является дисперсией воспроизводимости этого опыта, будем считать дисперсией воспроизводимости всего эксперимента.

Проверим нормальность распределения выходной величины $y(x_1, \dots, x_n)$ в фиксированной точке, т. е. нормальность распределения ϵ_i . Для этого у нас есть только 9-й опыт, который дублировался теперь уже 9 раз.

Воспользуемся критерием Колмогорова.

Оценка \tilde{a} математического ожидания и оценка s^2 дисперсии уже вычислены, $\tilde{a} = 13,0$; $s^2 = 0,0416$, следовательно, $s \simeq 0,204$.

Построим вариационный ряд $\{y_i^*\}$ по выборке $\{y_{9i}\}$:

I	1	2	3	4	5	6	7	8	9
y_i^*	12,7	12,8	12,9	12,9	12,96	13,1	13,1	13,2	13,35

Вычислим значения функции нормального распределения с параметрами (\tilde{a}, s) в точках вариационного ряда по формуле

$$z_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} e^{-\tau^2/2} d\tau = 0,5 + \Phi(t_i),$$

где $t_i = \frac{y_i^* - \tilde{a}}{s}$;

Составим таблицу:

I	1	2	3	4	5	6	7	8	9
t_i	-1,471	-0,980	-0,490	-0,490	-0,196	0,490	0,490	0,980	1,716
z_i	0,071	0,163	0,312	0,312	0,421	0,688	0,688	0,836	0,957

Вычислим значения $D(\tilde{X}_n)$: $d_i = \frac{i}{n} - z_i$, $i = 1, \dots, 9$, $d_{i+9} = z_i - \frac{i-1}{n}$, $i = 1, \dots, 9$.

Составим таблицу:

I	1	2	3	4	5	6	7	8	9
d_i	0,040	0,059	0,021	0,132	0,134	-0,021	0,090	0,053	0,043
d_{i+9}	0,071	0,052	0,090	-0,021	-0,023	0,132	0,021	0,058	0,068

Вычислим значение $d_{\text{расч}} = \max_{1 \leq i \leq 18} |d_i| = 0,134$ и для уровня значимости $q = 0,05$ найдем по табл. П2 приложения значение $d_{\text{табл}} = 0,43001$. Так как $d_{\text{расч}} < d_{\text{табл}}$, то гипотеза о нормальности распределения принимается.

Используя МНК, найдем Θ .

Составим матрицу спектра значений базисных функций:

$$\tilde{X} = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Матрица дублирования имеет вид

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 \end{pmatrix}.$$

Тогда

$$X^T X = \tilde{X}^T V \tilde{X} = \begin{pmatrix} 17 & 0 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 6 & 0 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 0 & 4 & 6 \end{pmatrix},$$

при этом $X^T Y = (221,83 \ 71,86 \ 84,12 \ -0,4 \ 80,54 \ 76,48)^T$.
Запишем систему уравнений

$$\begin{cases} 17\theta_0 + 6\theta_{11} + 6\theta_{22} = 221,83; \\ 6\theta_1 = 71,86; \\ 6\theta_2 = 84,12; \\ 4\theta_{12} = -0,4; \\ 6\theta_0 + 6\theta_{11} + 4\theta_{22} = 80,54; \\ 6\theta_0 + 4\theta_{11} + 6\theta_{22} = 76,48. \end{cases}$$

Решив систему, найдем

$$\begin{aligned} \theta_0 &= 13,02; \quad \theta_1 = 11,97; \quad \theta_2 = 14,02; \quad \theta_{12} = -0,1; \\ \theta_{11} &= 1,05; \quad \theta_{22} = -0,98. \end{aligned}$$

Проведем проверку значимости коэффициентов.

При вычислении коэффициентов не использовали обратную матрицу к матрице $X^T X$. Однако для проверки значимости коэффициентов ее необходимо вычислить, так как диагональные элементы этой матрицы c_{ii} нужны для вычисления дисперсии i -го коэффициента:

$$s^2 \{ \theta_i \} = c_{ii} s^2 \{ y \},$$

где $s^2 \{ y \}$ — дисперсия воспроизводимости;

$$(X^T X)^{-1} = \begin{pmatrix} \frac{5}{49} & 0 & 0 & 0 & \frac{-3}{49} & \frac{-3}{49} \\ 0 & \frac{1}{6} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{-3}{49} & 0 & 0 & 0 & \frac{33}{98} & \frac{-8}{49} \\ \frac{-3}{49} & 0 & 0 & 0 & \frac{-8}{49} & \frac{33}{98} \end{pmatrix}.$$

По уровню значимости $q = 0,05$ и числу степеней свободы, связанному с вычислением дисперсии воспроизводимости, $f_y = 8$, найдем в табл. П1 приложения значение $t_{\text{табл}} = 2,31$. Вычислим средние квадратичные отклонения коэффициентов и отбросим незначимые:

$$s(\theta_0) = \sqrt{\frac{5}{49}} \cdot 0,204 \cdot 2,31 \approx 0,151.$$

Так как

$$\theta_0 = 13,02 > s(\theta_0),$$

то θ_0 оставляем;

$$s(\theta_1) = \sqrt{\frac{1}{6}} \cdot 0,204 \cdot 2,31 \simeq 0,192.$$

Так как $\theta_1 = 11,97 > s(\theta_1)$, то θ_1 оставляем;

$$s(\theta_2) = \sqrt{\frac{1}{6}} \cdot 0,204 \cdot 2,31 \simeq 0,192.$$

Так как $\theta_2 = 14,02 > s(\theta_2)$, то θ_2 оставляем;

$$s(\theta_{12}) = \sqrt{\frac{1}{4}} \cdot 0,204 \cdot 2,31 \simeq 0,236.$$

Так как $|\theta_{12}| = 0,1 < s(\theta_{12})$, то θ_{12} — незначимый коэффициент, мы должны его отбросить;

$$s(\theta_{11}) = \sqrt{\frac{33}{98}} \cdot 0,204 \cdot 2,31 \approx 0,273.$$

Так как $\theta_{11} = 1,05 > 0,605$, то θ_{11} оставляем;

$$s(\theta_{22}) = \sqrt{\frac{33}{98}} \cdot 0,204 \cdot 2,31 \approx 0,273.$$

Так как $\theta_{22} = 0,98 > 0,605$, то θ_{22} оставляем.

Итак, получили модель:

$$y = 13,02 + 11,97x_1 + 14,02x_2 + 1,05x_1^2 - 0,98x_2^2.$$

Так как коэффициент θ_{12} оказался незначимым, то получилась модель нового типа:

$$y = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_{11}x_1^2 + \theta_{22}x_2^2. \quad (6)$$

Необходимо пересчитать коэффициенты.

Составим новую матрицу спектра базисных функций:

$$\tilde{X} = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Матрица дублирования V не меняется. Вычислим матрицы $X^T X$ и $X^T Y$:

$$X^T X = \tilde{X}^T V \tilde{X} = \begin{pmatrix} 17 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 6 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 4 & 6 \end{pmatrix}; \quad X^T Y = \begin{pmatrix} 221,83 \\ 71,86 \\ 84,12 \\ 80,54 \\ 76,48 \end{pmatrix}.$$

Составим систему уравнений:

$$\begin{cases} 17\theta_0 + 6\theta_{11} + 6\theta_{22} = 221,83; \\ 6\theta_1 = 71,86; \\ 6\theta_2 = 84,12; \\ 6\theta_0 + 6\theta_{11} + 4\theta_{22} = 80,54; \\ 6\theta_0 + 4\theta_{11} + 6\theta_{22} = 76,48. \end{cases}$$

Система для оставшихся коэффициентов оказалась прежней. Хотя матрица X для первоначальной модели (5) не является ортогональной, но у нее есть ортогональная часть. При изменении модели (5) исчезает столбец из ортогональной части, поэтому пересчет коэффициентов не требуется. При этом и дисперсии коэффициентов не изменяются, так как получаем

$$(X^T X)^{-1} = \begin{pmatrix} \frac{5}{49} & 0 & 0 & \frac{-3}{49} & \frac{-3}{49} \\ 0 & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 \\ \frac{-3}{49} & 0 & 0 & \frac{33}{98} & \frac{-8}{49} \\ \frac{-3}{49} & 0 & 0 & \frac{-8}{49} & \frac{33}{98} \end{pmatrix},$$

т. е. из прежней дисперсионной матрицы вычеркиваются строка и столбец, соответствующие отброшенному коэффициенту. Не следует забывать, что пересчета не требуется только потому, что отброшенный коэффициент относится к ортогональной части матрицы базисных функций.

Проведем проверку адекватности полученной модели (6). Так как коэффициентов всего 5, а опытов 9 и, кроме того, известна дисперсия воспроизводимости, то это возможно.

Вычислим значения \hat{y}_i выходной величины по модели (6) и квадраты отклонений вычисленных значений от результатов эксперимента и составим следующую таблицу:

I	1	2	3	4	5	6	7	8	9
\hat{y}_i	-12,9	15,14	11,04	39,08	2,1	26,04	-1,98	26,06	13,02
y_i	-12,9	15,2	11,1	38,8	2,04	26,3	-2,02	26,3	13,0
$(\hat{y}_i - y_i)^2$	0,0	0,0036	0,0036	0,0784	0,0036	0,0676	0,0016	0,0576	0,0004

Следовательно, $\sum_{i=1}^9 (\hat{y}_i - y_i)^2 = 0,2164$.

Число степеней свободы дисперсии адекватности $f_a = 9 - 5 = 4$.

Дисперсия адекватности $s_a^2 = \sum_{i=1}^9 (\hat{y}_i - y_i)^2 / 4 = 0,0541$.

Найдем значение $F_{\text{расч}} = s_a^2 / s^2\{y\} = \frac{0,0541}{0,0416} \simeq 1,30$.

По уровню значимости $q = 0,05$ и числу степеней свободы $k_1 = f_y = 8$ (число степеней свободы при вычислении дисперсии воспроизводимости) и $k_2 = f_a = 4$ найдем в табл. ПЗ приложения значение $F_{\text{табл}} = 6,04$.

Так как $F_{\text{расч}} < F_{\text{табл}}$, то модель (6) адекватна.

ПРИЛОЖЕНИЕ

Таблица П.1

**Критические точки распределения Стьюдента для двусторонней
критической области (q – уровень значимости;
 f – число степеней свободы)**

f	q			
	0,1	0,05	0,02	0,01
1	6,31	12,7	31,82	63,7
2	2,92	4,30	6,97	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,01	2,57	3,37	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	2,20	2,72	3,11
12	1,78	2,18	2,68	3,05
13	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98
15	1,75	2,13	2,60	2,95
16	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2,90
18	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86
20	1,73	2,09	2,53	2,85

Таблица П.2

Критические значения для наибольшего отклонения эмпирического распределения от теоретического (критерий Колмогорова) [4]

n	q				
	0,2	0,1	0,05	0,02	0,01
1	0,9	0,95	0,975	0,99	0,995
2	0,68377	0,77639	0,84189	0,9	0,92929
3	0,56481	0,63604	0,70760	0,78456	0,82900
4	0,49265	0,56522	0,62394	0,68887	0,73424
5	0,44698	0,50945	0,56328	0,62718	0,66853
6	0,41037	0,46799	0,51926	0,57741	0,61661
7	0,38148	0,43607	0,48342	0,53844	0,57581
8	0,35831	0,40962	0,45427	0,50654	0,54179
9	0,33910	0,38746	0,43001	0,47960	0,51332
10	0,32260	0,36866	0,40925	0,45662	0,48893
11	0,30829	0,35242	0,39122	0,43670	0,46770
12	0,29577	0,33815	0,37543	0,41918	0,44905
13	0,28470	0,32549	0,36143	0,40362	0,43247
14	0,27481	0,31417	0,34890	0,38970	0,41762
15	0,26588	0,30397	0,33760	0,37713	0,40420
16	0,25778	0,29472	0,32733	0,36571	0,39201
17	0,25039	0,28627	0,31796	0,35528	0,38086
18	0,24360	0,27851	0,30936	0,34569	0,37062
19	0,23735	0,27136	0,30143	0,33685	0,36117
20	0,23156	0,26473	0,29408	0,32866	0,35241
21	0,22617	0,25858	0,28724	0,32104	0,34427
22	0,22115	0,25283	0,28087	0,31394	0,33666
23	0,21645	0,24746	0,27490	0,30728	0,32954
24	0,21205	0,24242	0,26931	0,30104	0,32286
25	0,20790	0,23768	0,26404	0,29516	0,31657
26	0,20399	0,23320	0,25907	0,28962	0,31064
27	0,20030	0,22898	0,25438	0,28438	0,30502

Окончание табл. П.2

n	q				
	0,2	0,1	0,05	0,02	0,01
28	0,19680	0,22497	0,24993	0,27942	0,29971
29	0,19348	0,22117	0,24571	0,27471	0,29466
30	0,19032	0,21756	0,24170	0,27023	0,28987
31	0,18732	0,21412	0,23788	0,26596	0,28530
32	0,18445	0,21085	0,23424	0,26189	0,28094
33	0,18171	0,20771	0,23076	0,25801	0,27677
34	0,17909	0,20472	0,22743	0,25429	0,27279
35	0,17659	0,20185	0,22425	0,25073	0,26897
36	0,17418	0,19910	0,22119	0,24732	0,26532
37	0,17188	0,19646	0,21826	0,24404	0,26180
38	0,16966	0,19392	0,21544	0,24089	0,25843
39	0,16753	0,19148	0,21273	0,23786	0,25518
40	0,16547	0,18913	0,21012	0,23494	0,25205
41	0,16349	0,18687	0,20760	0,23213	0,24904
42	0,16158	0,18468	0,20517	0,22941	0,24613
43	0,15974	0,18257	0,20283	0,22679	0,24332
44	0,15796	0,18053	0,20056	0,22426	0,24060
45	0,15623	0,17856	0,19837	0,22181	0,23798
46	0,15457	0,17665	0,19625	0,21944	0,23544
47	0,15295	0,17481	0,19420	0,21715	0,23298
48	0,15139	0,17302	0,19221	0,21493	0,23059
49	0,14987	0,17128	0,19028	0,21277	0,22828
50	0,14840	0,16959	0,18841	0,21068	0,22604

Таблица П.3

Критические точки распределения F Фишера – Снедекора
(f_1 – число степеней свободы дисперсии числителя, f_2 – число
степеней свободы дисперсии знаменателя, уровень значимости
 $q = 0,05$)

f_2	f_1									
	1	2	3	4	5	6	7	8	9	10
1	161,5	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54

ЛИТЕРАТУРА

1. *Налимов В.В.* Статистические методы планирования экстремальных экспериментов / В.В. Налимов, Н.А. Чернова. М.: Наука, 1965.
2. *Горяинов В.Б.* Математическая статистика / В.Б. Горяинов, И.В. Павлов, Г.М. Цветкова и др. ; под ред. В.С. Зарубина и А.П. Крищенко. М.: Изд-во МГТУ им. Н.Э. Баумана, 2001.
3. *Драйпер Н.* Прикладной регрессионный анализ : пер с англ. / Н. Драйпер, Г. Смит. М.: Статистика, 1973.
4. *Большев Л.Н.* Таблицы математической статистики / Л.Н. Большев, Н.В. Смирнов. М.: Наука, 1983.
5. *Адлер Ю.П.* Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер, Е.П. Маркова, Ю.В. Грановский. М.: Наука, 1976.
6. *Вентцель Е.С.* Теория вероятностей / Е.С. Вентцель. М.: Наука, 1969.
7. *Гмурман В.Е.* Теория вероятностей и математическая статистика / В.Е. Гмурман. М.: Высш. шк., 1999.
8. *Грешилов А.А.* Математические методы принятия решений / А.А. Грешилов. М.: Изд-во МГТУ им. Н.Э. Баумана, 2006.
9. Математическая теория планирования эксперимента / под ред. С.М. Ермакова. М.: Наука, 1983.
10. Математические методы планирования эксперимента / под ред. В.В. Пененко. Новосибирск: Наука, 1981.
11. *Себер Дж.* Линейный регрессионный анализ : пер с англ. / Дж. Себер. М.: Мир, 1980.
12. *Асатурян В.И.* Методические указания к домашнему заданию по курсу «Теория планирования эксперимента» / В.И. Асатурян, Н.Т. Вилисова. М.: Изд-во МГТУ им. Н.Э. Баумана, 1983.

ОГЛАВЛЕНИЕ

Введение	3
1. Метод наименьших квадратов	6
2. Оценивание параметров модели при повторных наблюдениях ...	10
3. Отбрасывание грубых наблюдений	13
4. Статистический анализ уравнения регрессии	13
5. Проверка адекватности регрессионной модели	17
6. Домашнее задание по математическому моделированию и плани- рованию эксперимента	19
7. Решение типового примера	19
Приложение	29
Литература	33

Учебное издание

Полякова Надия Салихжановна
Дерябина Галина Сергеевна
Федорчук Хелина Романовна

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА**

Редактор *О.М. Королева*
Корректор *М.А. Василевская*
Компьютерная верстка *В.И. Товстоног*

Подписано в печать 20.10.2010. Формат 60×84/16.
Усл. печ. л. 2,09. Тираж 300 экз. Изд. № 3.
Заказ

Издательство МГТУ им. Н.Э. Баумана.
Типография МГТУ им. Н.Э. Баумана.
105005, Москва, 2-я Бауманская ул., 5.

ДЛЯ ЗАМЕТОК